МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

А.И. Данилин

Основы теории оптимизации (постановки задач)

Электронное учебное пособие

САМАРА
2011

УДК 51.380.115
ББК 22.18
Д 182

Данилин А. И., Основы теории оптимизации (постановки задач) [Электронный ресурс] : электрон. учеб. пособие / А. И. Данилин; Минобрнауки России, Самар. гос. аэрокосм. ун-т им. С. П. Королева (нац. исслед. ун-т). - Электрон. текстовые и граф. дан. (1,2 МБайт). - Самара, 2011. - 1 эл. опт. диск (CD-ROM).

Рассмотрены базовые идеи и методы постановок задач оптимизации параметров в организационно-технических системах. Основное внимание уделено понятийным аспектам теории оптимизации, которые встречаются при оптимизации организационных и технических систем, агрегатов, технологических процессов, а также при принятии решений.

Учебное пособие предназначено для студентов факультета инженеров воздушного транспорта, обучающихся по направлению подготовки магистров специальностей 162300.68 "Техническая эксплуатация летательных аппаратов и двигателей" и 162500.68 «Техническая эксплуатация авиационных электросистем и пилотажно-навигационных комплексов», дисциплина «Методы оптимизации», семестр 9.

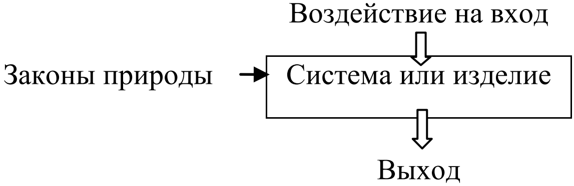
Разработано на кафедре эксплуатации авиационной техники.

1. Введение.

На протяжении всей своей истории люди при необходимости принимать решения прибегали к сложным ритуалам. Они устраивали торжественные церемонии, приносили в жертву животных, гадали по звездам и следили за полетом птиц. Они полагались на народные приметы и старались следовать примитивным правилам, облегчающим им трудную задачу принятия решений. Ведь цена ошибки зачастую была очень велика. Так появилась астрология, нумерология, хиромантия и другие оккультные учения.

В настоящее время для принятия решения используют новый и, по-видимому, более научный «ритуал», основанный на применении компьютера. Без современных технических средств человеческий ум, вероятно, не может учесть многочисленные и разнообразные факторы и их связи, с которыми сталкиваются при управлении авиационным предприятием, конструировании, изготовлении и испытании воздушного судна или оптимизации регламентных работ. Существующие в настоящее время многочисленные методы ***оптимизации*** уже достаточно развиты, что позволяет эффективно использовать возможности цифровых и гибридных вычислительных машин.

Вообще, всю нашу осмысленную деятельность в любой сфере можно пояснить следующей простой схемой.



Дано	Найти	Название процесса
Система или изделие, воздействие на вход, законы природы	Выход	Анализ (прогнозирование и/или расчёт характеристик, поведения, последствий… во всех сферах деятельности)
Система или изделие, выход, законы природы	Воздействие на вход	Инверсный анализ (медицина, театр, расследование происшествий, планирование деятельности…)
Система или изделие, воздействие на вход, выход	Законы природы	Научные исследования
Воздействие на ход, выход, законы природы	Систему или изделие	Инженерное проектирование

Инженерное проектирование является на сегодняшний день интеллектуально самым трудным видом деятельности вне зависимости от объекта проектирования. В названии деятельности термин «инженерный» используется в его изначальном смысле, а именно: ingenium в переводе с латыни означает изобретение; то есть, «инженерный» означает «творческий, с элементами изобретательства», а инженер – это изобретатель. Равно как и термин «проект» происходит от латинского proectus – брошенный вперед; то есть, проект – это нечто такое, чего раньше не существовало.

Методологические аспекты проектирования привлекают внимание многих исследователей, поскольку осмысление этого процесса и выявление его закономерностей, возможно, дадут повышение эффективности результатов проектирования, каковыми являются организационно-технические системы и/или изделия. К сожалению, успехов в оптимизации самого процесса проектирования немного, однако сформулировано уже более двух десятков определений проектирования, из которых приведём только четыре.

1. Проектирование – это принятие решения в условиях неопределённости с тяжелыми последствиями в случае ошибки;- автор: Азимов.

2. Проектирование – это приведение изделия (системы) в соответствие с обстановкой при максимальном учёте всех требований; - автор: Грегори.

3. Проектирование – это вдохновенный прыжок от фактов настоящего к возможностям будущего; - автор: Пейдж.

4. Проектирование – это использование научных принципов, технической информации и воображения для определения … структуры машины или системы, предназначенной для выполнения заранее заданных функций с наибольшей экономичностью и эффективностью; - автор: Филден.

Эти определения неявно предполагают три вещи.

1. Цель проектирования известна, то есть мы знаем, по крайней мере, функции, которые должен выполнять проектируемый технический объект или система.

2. Эти функции невозможно удовлетворительно решить с помощью уже существующих технических объектов или систем.

3. Создаваемый новый объект или система должны быть оптимальными, то есть функции должны выполняться с максимально достижимой эффективностью.

На рис. 1 показаны этапы создания новой системы.

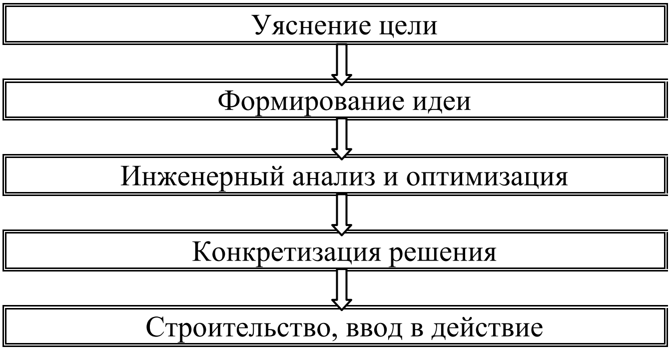


Рис. 1. Этапы создания системы или изделия

При решении организационно-технической задачи обычно бывает необходимо решить два ключевых вопроса.

1. Инженер должен предложить (выдумать, разработать) метод, схему или идею, которые лягут в основу функционирования новой системы и, по его мнению, отвечают поставленным требованиям.

2. Затем инженер должен *количественно* проанализировать свой метод, схему или идею, с тем, чтобы убедиться, что они могут удовлетворить поставленным требованиям при заданных ограничениях.

После положительных результатов второго этапа необходимо попытаться улучшить характеристики проектируемой системы, и здесь без оптимизации добиться желаемой эффективности может только гениальный или невероятно удачливый проектировщик.

Что же это такое - оптимальный проект, оптимальная система? Как оптимальность соотносится с другими свойствами изделия или организационно-технической системы, если таковые имеются?

Термин "оптимальный" сильно перегружен смысловым наполнением. В переводе с латинского он означает "наилучший" и корнями уходит к богине **Опе** - богине плодородия и урожая у древнеиталийского племени сабинов [1]. Впоследствии богиня Опа стала женой бога времени Сатурна и матерью Юпитера. Она изображается держащей в одной руке рог изобилия мифических благ, а в другой весы, символизирующие измерение и принятие решения.

Когда употребляется слово "*оптимальный*", то в нем всегда подразумевается критерий, по которому данное техническое решение является наилучшим. Этот критерий может быть ***единственным свойством*** системы, а может быть ***целым комплексом свойств***. И оптимальный означает наилучший по выбранному критерию объект.

Оптимальный по эффективности двигатель имеет (в идеале) коэффициент полезного действия равный единице; оптимальный идеальный самолет не требует затрат на обслуживание и переносит пассажиров по воздуху за нулевое время или с нулевыми затратами топлива в зависимости от принятого критерия; оптимальный выпускник Самарского государственного аэрокосмического университета трудные задачи решает немедленно, а невозможные - немного погодя.

В настоящее время методы оптимизации разделились на два больших класса: инженерные методы и методы математического программирования.

Инженерные методы оптимизации, или как их ещё называют, методы критериев оптимальности, имеют объектом проектирования в основном материальные изделия, которые обладают набором физических свойств, не зависящих от человеческих факторов. В них выявляются свойства оптимальных проектов, так называемые критерии оптимальности, и строятся итерационные процедуры направленного изменения параметров объекта, приводящие к достижению критериев оптимальности. При этом выполнение критериев оптимальности обеспечивает оптимальность объекта в целом.

Другой большой класс методов оптимизации представляют методы математического программирования. Они обладают необходимой общностью и применимы к любым организационно-техническим задачам.

Название «математическое программирование» предложено Робертом Дорфманом приблизительно в 1950г.; теперь оно объединяет линейное программирование, целочисленное программирование, динамическое программирование, выпуклое программирование, нелинейное программирование и программирование при наличии неопределенности. Причём «программирование» здесь не имеет ничего общего с реальным написанием кода программ, а должно читаться по сути своего содержания как «методы оптимизации».

Для технических задач наиболее применимо нелинейное программирование, которое имеет дело с оптимизацией нелинейных функций при линейных и/или нелинейных ограничениях. Типичными областями его применения являются прогнозирование, планирование промышленного производства, управление материальными ресурсами, контроль качества выпускаемой продукции, планирование обслуживания и ремонта, проектирование технологических линий (процессов), учет и планирование капиталовложений, а также проектирование агрегатов и технических систем. Пока еще не существует универсального метода решения нелинейных задач оптимизации, такого, как, например, симплексный алгоритм, разработанный для задач линейного программирования. Нелинейное программирование при решении задач включает в себя элементы экспериментирования и подбора наиболее эффективной процедуры отыскания экстремума, учитывающей специфику решаемой задачи.

2. Постановка задачи оптимизации.

При решении любой организационной или технической задачи её сначала необходимо сформулировать, то есть ответить на вопросы: 1) *С чем мы имеем дело?* Каковы функции системы, каковы её связи с внешним окружением? 2) *Чего мы хотим?* Что нужно сделать, или как нужно сделать? 3) *Каковы наши возможности*? Какие у нас есть ресурсы для исполнения наших желаний?

Именно здесь совершается много ошибок. Как точно заметил Джон Дьюи: «Хорошо поставленная задача уже наполовину решена». Задача, сведённая к конкретным вопросам, позволяет выразить получаемое решение через величины, которые можно затем вычислить или измерить. Другими словами, нужно ставить такие вопросы, на которые можно получить *количественный* ответ. Недостаточно ставить вопросы типа: Будет ли самолёт экономичным? Будет ли конструкция работать? Такие вопросы не являются конкретными. Вместо этого нужно ставить вопросы иного характера. Какова будет стоимость тонно-километра при перевозке 30 пассажиров на дальность 1000 км при рейсовой скорости 450 км/час? Какие напряжения возникнут в элементах конструкции при нагрузке 80 кН и нагреве до 300^0C? Для подготовки таких вопросов, то есть для определения задачи необходимо чёткое понимание реальной физической ситуации и умение рассуждать. Формулировке задачи обычно мешают избыточные условия, которые делают технические задачи неимоверно сложными. И здесь огромную помощь оказывает определение задачи в терминах математического программирования. Согласно ей, при постановке задачи математического программирования необходимо выполнить следующие работы: 1) выбрать функцию цели; 2) выбрать проектные переменные; 3) записать ограничения.

Задача оптимизации формулируется так: найти вектор проектных переменных $\vec{X}(x_1, x_2, ... x_n)$***, обеспечивающий минимальное значение целевой функции*** $f(\vec{X}) \Rightarrow \min$ ***при выполнении системы функциональных и геометрических неравенств:*** $g_j(\vec{X}) \le 0;\ j = 1,2,..m;\ a_i \le x_i \le b_i;\ i = 1,2,..n$.

2.1. Функция цели.

Это критерий качества изделия, процесса или системы, минимума или максимума которого мы хотим достичь. Численное значение функции цели определяет степень совершенства объекта, разработанного в результате проектирования. Функция цели должна быть *одна* и в этом состоит трудность её выбора.

Нельзя создать самолёт, обладающий максимальной скоростью, высотой и дальностью полёта; нельзя сделать шасси с минимальной массой и максимальной энергоёмкостью; нельзя создать воздушное судно с минимальными затратами на техническое обслуживание и максимальной надёжностью.

Для организационно-технической системы, например при проектировании авиационно-технологической базы, различными руководителями среднего звена могут быть выбраны разные критерии качества системы. В частности:

1. Начальник отдела снабжения отдаст предпочтение проекту, согласно которому производственный процесс протекает с минимальными изменениями в ассортименте расходных материалов и запасных частей. При этом минимизируются затраты по доставке и входному контролю.

2. Руководитель производственного отдела заинтересован в реализации проекта, в соответствии с которым максимальны объемы складских запасов расходных материалов и запасных частей всех видов. В этом случае минимизируется продолжительность интервала времени обслуживания и ремонта воздушных судов между получением заказа от потребителя и его выполнением.

3. Руководитель аэропорта отдаст предпочтение проекту, в соответствии с которым объемы складских запасов минимальны, с тем, чтобы уменьшить капитальные издержки на хранение изделий.

Нетрудно видеть, что эти критерии при оптимизации не могут быть реализованы одновременно.

Здесь мы подошли к вопросу: как быть, если к проекту предъявляются минимаксные требования по многим критериям? Подробные и предельно чёткие ответы на этот вопрос можно найти в книге Д.И.Батищева [2], но из всех рекомендаций по построению единственной функции цели отметим два подхода, чтобы проиллюстрировать общее направление мыслей.

1. Метод взвешенных сумм. Функция цели имеет вид: $F(\overrightarrow{X}) = \sum_{k=1}^{l} \lambda_k f_k(\overrightarrow{X})$, где $f_k(\overrightarrow{X})$ - частные критерии качества; λ_k – весовые коэффициенты, показывающие *важность* отдельного критерия в общей оценке качества проекта.

2. Оценка по потерям. Функция цели имеет вид: $F(\overrightarrow{X}) = \sum_{k=1}^{l} \lambda_k \left[f_k(\overrightarrow{X}) - f_k^*(\overrightarrow{X}) \right]$. Здесь $f_k^*(\overrightarrow{X})$ - минимальное или максимальное значение частного критерия качества.

Основная трудность в реализации этих подходов состоит в назначении весовых коэффициентов λ_k. От их выбора зависит облик будущего изделия или системы и все их параметры и характеристики. Отметим ещё, что при проектировании одного и того же объекта весовые коэффициенты λ_k в первом и втором подходе могут быть (и зачастую бывают) различными.

2.2. Проектные переменные.

Это независимые параметры изделия, процесса, системы, различным численным значениям которых соответствуют различные характеристики проектируемого изделия, процесса или системы. Проектные переменные независимы в абсолютном смысле, - это те изначальные величины, вариацией которых добиваются заданных функциональных характеристик объекта проектирования и минимума функции цели.

Выбор проектных переменных – дело далеко не тривиальное. Набор проектных переменных должен позволять получать в процессе оптимизации *новые технические решения*, а не только улучшать существующие объекты. На эту сторону проектных переменных обычно не обращают внимания, искусственно сужая возможности оптимизации.

Например, при модификации самолёта было принято решение изменить аэродинамический профиль крыла. Какой профиль выбрать по критерию максимального аэродинамического качества?

Можно обратиться к доступному сортаменту профилей и выбрать какой-либо профиль из них. А может быть есть другие профили? Конечно, есть и вопрос об оптимальности принятого решения не закрыт.

Попытаемся ввести более универсальные проектные переменные. Для этого выберем в качестве проектных переменных координаты конечного множества точек в какой-либо системе координат $\xi\eta$ и будем считать, что профиль представляет собой замкнутую линию, проведённую через эти точки, см. рис. 2. Нетрудно видеть, что различным определённым значениям этих проектных переменных могут соответствовать любые профили из любых сортаментов и множество новых.

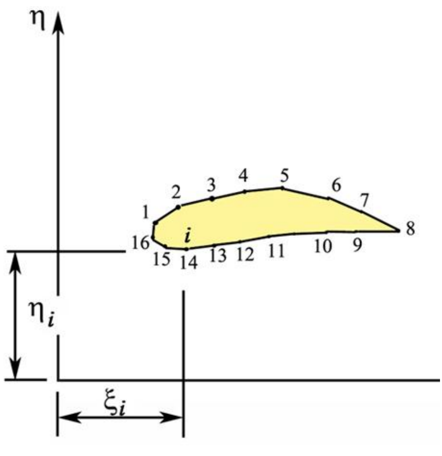


Рис. 2. Задание профиля крыла.

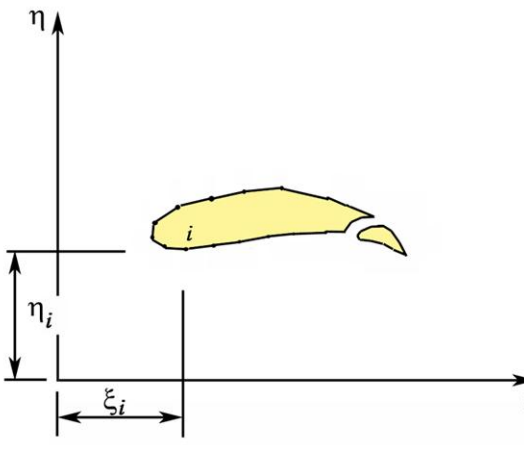


Рис. 3. Двухсвязная область

Такое пространство проектных переменных значительно богаче возможностями, то есть оно более универсально. Однако и у него есть недостаток. При таком описании сечения мы не сможем получить двухсвязную область, например, профиль крыла с закрылком, см. рис. 3.

Этот пример показывает, насколько важно и сложно выбрать достаточно универсальный набор проектных переменных, который, в принципе, должен приводить к новым техническим решениям. При этом инженер обязан помнить, что дом из кирпичей не обязательно имеет форму кирпича.

Ещё одним существенным фактором, влияющим на выбор проектных переменных, является уровень детализации при исследовании системы. Очень важно ввести в рассмотрение все основные независимые переменные, но не менее важно не «перегружать» задачу большим количеством мелких, несущественных деталей. Например, при проведении предварительного анализа производственного процесса, включающего большое число различных видов оборудования - стендов, станков, насосов, компрессоров, теплообменников, - обычно явным образом не учитываются несущественные особенности конструкции того или иного устройства. Насос может быть с достаточной для поставленных целей полнотой охарактеризован его производительностью, энергопотреблением и перепадами давления в трубах и корпусе аппарата. Переменные, имеющие непосредственное отношение к конструкции аппарата, - количество и размеры труб, фланцы подсоединения магистралей, размеры корпуса - рассматриваются, как правило, при исследовании каждого отдельного устройства, но в технической системе представлены своими характеристиками, которые являются варьируемыми параметрами. При выборе проектных переменных целесообразно руководствоваться правилом, согласно которому следует рассматривать только те переменные, которые оказывают существенное влияние на функцию цели, выбранную для анализа сложной системы.

Далее важно провести различие между теми параметрами системы, которые могут предполагаться детерминированными, и параметрами, которые подвержены флуктуациям вследствие случайного воздействия внешних или неконтролируемых факторов. В частности, в примере с авиационно-технологической базой поломки оборудования и прогулы рабочих могут оказать весьма заметное влияние на длительность процесса обслуживания самолётов. Ясно, что существенные изменения этих важных параметров системы должны быть приняты во внимание при постановке задачи производственного планирования, если требуется, чтобы составленный оптимальный план был реалистичным и выполнимым.

2.3. Запись ограничений.

Ограничения бывают двух типов. Во-первых, это ограничения на диапазон изменения самих проектных переменных: $a_i \leq x_i \leq b_i$, $i=1,2,..n$. Такие ограничения называются геометрическими. Например, толщина обшивки может быть ограничена конструктивно: $\delta_{min} \leq \delta_i \leq \delta_{max}$.

Вторым видом ограничений являются функциональные ограничения, связанные с работоспособностью проектируемой системы. На словах они часто выражаются довольно просто, и на первый взгляд кажется, что это просто сделать и математически. Например, «конструкция должна иметь достаточную прочность и долговечность». Это требование включает в себя вычисление напряжений во всех расчётных случаях нагружения (до 36) для каждого элемента конструкции, что предполагает построение адекватной модели конструкции, моделировании нагрузок и проведение вычислений. Всё это требует значительных затрат календарного времени и высокой квалификации инженера-расчётчика.

Ограничения обычно записываются в стандартной форме $g_j(\overrightarrow{X}) \geq 0;\ j=1,2,..m$. Они выделяют из всего множества проектных переменных подмножество, являющееся областью допустимых проектов. Главная неприятная особенность функциональных ограничений в технических задачах – это то, что они носят алгоритмический характер. Мы не можем по значениям проектных переменных сразу определить, принадлежит ли этот набор параметров области допустимых проектов. Мы должны рассчитать систему с данными значениями проектных переменных, определить её характеристики и сравнить с допустимыми.

Здесь можно провести наглядную аналогию с задачей отыскания наиболее глубокого места озера, но только если дно в этой точке свободно от водорослей. Чтобы замерить глубину и чистоту дна нужно подплыть на лодке в заданную точку, опустить на дно драгу для добывания водорослей и только после того, как драга ничего не принесла, - замерять глубину.

Функциональные и геометрические ограничения могут привести к следующим ситуациям.

1. Допустимая область проектных решений является пустой.

2. Допустимая область проектных решений содержит единственную точку.

3. Допустимая область содержит множество проектных решений.

В первом случае в рамках наложенных ограничений проект реализовать невозможно. Во втором случае проект единственный и ни о какой оптимизации речи не идёт. В третьем случае возможно несколько проектных решений и мы должны выбрать наилучшее, то есть провести оптимизацию.

3. Построение модели.

Модель является непременным атрибутом оптимизации, поскольку она связывает проектные переменные, характеристики проектируемого объекта и сам объект в процессе функционирования.

Инженерные модели не являются точными копиями реальных объектов, явлений и процессов. Реальность очень сложна и её никогда нельзя смоделировать в полном объёме. Всегда делают допущения и пользуются аппроксимацией. Построение воображаемой модели реального явления – это обычная процедура в инженерной практике. Важно только, чтобы инженер *знал о различии* между своей моделью и реальным миром. Эти различия определяются допущениями, независимо от того, приняты они сознательно или нет.

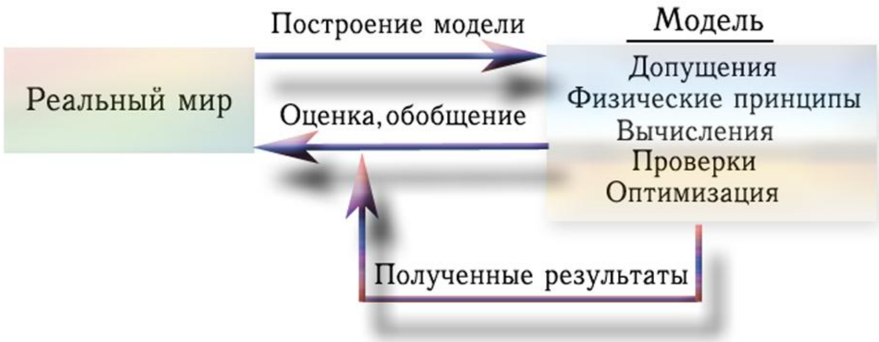


Рис. 4. Построение модели объекта.

К математическим моделям предъявляются следующие требования.

1. Адекватность модели – возможность получения имеющих смысл результатов.

2. Точность модели – согласованность расчётных данных с натурным экспериментом.

3. Быстродействие модели – возможность получения расчётных данных за приемлемое, обычно заданное, календарное время.

4. Работоспособность во всём диапазоне изменения проектных переменных.

Требования (1,2) и 3 противоречивы и поэтому процесс построения модели всегда связан с компромиссом. С одной стороны, адекватность и точность требуют детальных и сложных моделей, учитывающих все переменные, тогда как с другой стороны, время проектирования ограниченно и результат должен быть получен к строго заданному сроку.

3.1. Принятие допущений.

Какие именно допущения следует принимать при построении модели, в значительной мере зависит от целевой функции, учитываемых ограничений, и, вообще, от целей исследования.

Если нам нужно знать лишь порядок расчётной величины, то принимается одна группа допущений. Если же требуется погрешность менее 1%, то необходима другая группа допущений. Если для получения ответа отводится один час, то принимаются совершенно иные допущения, чем в том случае, когда на решение задачи даётся две недели.

Серьёзную опасность представляет принятие допущений без должного их понимания. Это недопустимо, поскольку инженер в этом случае не видит границ применимости своей модели. Всегда, начиная и заканчивая анализ, необходимо знать о принятых допущениях, включая *скрытые допущения*, которые прячутся, как правило, в формулах и методиках, взятых из справочной литературы. Технические справочники набиты формулами и методиками, но они не сообщают, при каких допущениях получены формулы, какие факторы учтены при их выводе, каковы границы их применимости. К примеру, если построить модель расчёта переходных процессов в полупроводниковых системах и забыть, что она верна только при определённых напряжениях и токах, то можно подать на вход такой модели напряжение 30 000 В и она сосчитает все переходные процессы, а также построит их графики. Что же будет в реальности? Легкий дымок и всё…

3.1.1. Некоторые часто используемые допущения.

1. Упругая балка. Когда удовлетворяются допущения простейшей теории упругой балки? Когда сечения плоские до деформаций остаются плоскими после деформаций? Прежде всего тогда, когда деформации являются упругими! Затем, когда строительная высота мала по сравнению с её длиной. Насколько мала? В соотношении 1:100 и менее. А если это соотношение будет 1:50 или 1:20 или 1:10? Здесь уже нужно учитывать, с какой точностью необходим результат. Кстати, фюзеляжи современных самолётов в силу наличия большого количества подкреплённых вырезов не подчиняются теории упругой балки, хотя часто рассчитываются именно по ней. Дело в том, что неточности расчёта поглощаются избыточным запасом прочности, связанным с

ограничениями на минимальные толщины конструктивных элементов.

2. Идеальный газ. Разумеется это только модель. Идеальный газ подчиняется уравнению $p\ v\ =\ R\ T$, в нём отсутствует внутреннее трение. В природе идеального газа нет. Обычным показателем того, что поведение газа близко к идеальному, является то, что давление p низко, а температура T высока по сравнению с критическими значениями. Понятие идеального газа используется в аэродинамике, закладывается в расчётные формулы аэронавигационных приборов.

3. Ньютоновская жидкость. Напряжение в жидкости пропорционально скорости деформаций. Коэффициент пропорциональности – вязкость. Вязкость – функция температуры и в меньшей мере давления. При больших скоростях деформаций (при ударе) это допущение неправомочно. Кроме того, нужно учитывать и другие свойства жидкости, - так, например, клей – не ньютоновская жидкость. Данное допущение используется в расчётах гидравлических агрегатов и процессов.

4. Адиабатная стенка. Отсутствует теплообмен. Допущение верно, если имеется теплоизоляция, если рассматривается стационарная задача, если время процесса очень мало. Например, сжатие газа в модели жидкостно-газового амортизатора шасси происходит по ударной адиабате без учёта теплообмена с внешней средой.

5. Сосредоточенные параметры. Такие допущения наиболее часто используются в теоретической механике, динамике полёта и электротехнике. При исследовании движения тела его масса помещается в центр масс и дополняется моментами инерции; электрическое сопротивление проводника, индуктивность катушки и ёмкость конденсатора считаются помещёнными в какие-то точечные позиции электрической схемы. В большинстве случаев не представляет труда определить эффект от допущения о сосредоточенных параметрах.

6. Свойства констант. Всегда ли постоянны физические константы? Модуль упругости материалов зависит от скорости

деформаций; плотность – функция температуры и давления; теплопроводность и удельная теплоёмкость – функции температуры; вязкость – функция температуры и давления; и т.д. Всё это надо учитывать при разработке конкретной модели.

7. Свойства окружающей среды. Всегда ли постоянны нормативные акты, процентные ставки кредитов, законы, регламентирующие авиатранспортную деятельность, наставления по производству полётов и т.д. Если мы хотим получить оптимальную и одновременно устойчивую к внешним воздействиям организационно-техническую систему, то необходимо проработать и возможные изменения окружающих условий.

Всегда нужно задавать себе вопросы: Даёт ли модель имеющие смысл результаты, если все входные параметры устремить к каким-либо пределам из допустимых диапазонов? Дополняют ли друг друга факторы, которые должны оказывать одинаковое воздействие на выходную характеристику? Все ли существенные факторы учтены? Не нарушает ли полученный результат физических законов?

3.2. Использование физических принципов.

Количественные и качественные связи между различными параметрами и характеристиками объекта проектирования всегда базируются на физических законах, которые в зависимости от принятых допущений и системы координат преломляются в различные уравнения. Так, закон сохранения энергии в аэродинамике представлен уравнением Бернулли для идеального несжимаемого газа и уравнениями Навье-Стокса для вязкого сжимаемого газа; в механике деформируемого твёрдого тела этот закон трансформируется в уравнение Лагранжа и две теоремы Кастильяно; в электротехнике – в уравнения Кирхгофа; в термодинамике и теплопередаче – в начала термодинамики, - список можно продолжать.

Во всех случаях при выявлении связей между параметрами и характеристиками лучше всего использовать физические принципы в их наиболее простом, ясном и фундаментальном виде.

Основных физических принципов немного, перечислим их.

1. Закон сохранения массы ⇨ аэродинамика, ракетная механика, химия, механика жидкости и газа, гидравлика…

2. Закон сохранения энергии ⇨ используется везде !

3. Законы сохранения движения ⇨ теоретическая механика, теория газотурбинных двигателей, теория вертолёта, механика тел переменной массы…

4. Законы Ньютона ⇨ техническая механика.

5. Эмпирические законы в механике: закон Гука, теории прочности материала, критерии подобия, законы трибологии…

6. Эмпирические законы электричества и магнетизма: законы Ома, Фарадея, Джоуля-Ленца, Кулона, Максвелла…

7. Эмпирические законы, регулирующие передачу тепла: лучистый и конвективный теплообмен, теплопроводность.

Так вот, при использовании какого-либо уравнения в модели, всегда следует отдавать себе отчёт в том, какой физический принцип оно выражает. Если же приходится самому выводить расчётные формулы и уравнения, то нужно использовать самый общий принцип.

3.3. Вычислительные трудности.

Математическая модель содержит функции, участвующие в процедуре оптимизации. Очевидно, для того чтобы искомый экстремум имел физический смысл, выбранная модель должна адекватно отражать существенные черты реального процесса. Но даже если это требование выполнено, при построении модели встречаются следующие типичные затруднения:

1. Функция цели может быть *нечувствительной* к изменениям некоторых проектных переменных и поэтому не удается определить четко выраженный экстремум.

2. Функция цели или некоторые из ограничений могут принимать в области поиска решения неограниченные значения; значения частных производных в математической модели также могут стать неограниченными. Особенно подвержены этой опасности модели с полиномами в знаменателе. Так, например, значения функции: $y = \dfrac{b_0 + b_1x_1}{b_2x_1 + b_3x_2}$; и ее первой частной производной по x_1:

$$\frac{\partial y}{\partial x_1} = \frac{-b_0b_2 + b_1b_3x_2}{\left(b_2x_1 + b_3x_2\right)^2};$$

обращаются в бесконечность при $b_2\ x_1 = -\ b_3\ x_2$. Эту трудность можно преодолеть, ограничив области допустимых значений проектных переменных путем введения дополнительных ограничений в задачу, или же изменив формулировку самой математической модели.

3. Переменные могут быть плохо *масштабированы*. Трудности масштабирования могут возникнуть, например, когда один из членов в выражении для критерия имеет существенно иной порядок величины, чем другой. При этом критерий становится нечувствительным к изменениям значений переменных в меньшем члене. Например, значение целевой функции $y = 100x_1^2 - 0.01x_2^2$ будет мало зависеть от изменения x_2, если только значение x_2 не окажется много больше значения x_1. Если x_2 представляет собой величину того же порядка, что и x_1, то либо одну переменную, либо обе переменные можно умножить на масштабные множители, в результате чего оба члена в правой части приведенного выше соотношения окажутся величинами одного и того же порядка. Положим, например,

$$\tilde{x}_1 = 10x_1, \qquad x_1^2 = 10^{-2}\tilde{x}_1^2,$$
$$\tilde{x}_2 = 10^{-1}x_2, \qquad x_2^2 = 10^{2}\tilde{x}_2^2.$$

Тогда члены в выражении для целевой функции становятся величинами одного порядка. После того как найден экстремум для

$y = \tilde{x}_1^2 - \tilde{x}_2^2$, можно определить значения x_1 и x_2 по значениям $\tilde{x}_1$ и $\tilde{x}_2$.Конечно, не всегда удаётся так легко изменить масштаб функций математической модели, как это сделано в рассмотренном примере.

Плохое масштабирование проектных переменных возможно также в случаях, если ограничения и/или целевая функция зависят от них в разной степени. Например, при оптимизации подкреплённых панелей масса панели зависит от толщины обшивки линейно, а от углов ориентации подкрепляющих наборов через $\sin^4\alpha$ и $\cos^4\alpha$, где α – угол ориентации набора. Изменение на один градус угла ориентации набора приведет к изменению целевой функции порядка $10^{-7}\rho S$, а изменение толщины обшивки на один миллиметр – порядка ($1\ \rho S$). Здесь ρ – плотность материала панели, а S – её площадь в плане. В этом случае масштабирование переменных требует изобретательности и глубокого понимания механики анизотропных пластин.

4. В плохо построенной математической модели переменные могут оказаться *взаимосвязанными*. Взаимное влияние переменных можно проиллюстрировать на примере очень простого критерия, в который входит произведение двух параметров: $y = 2\,x_1 x_2 + 10$. Здесь каждая из переменных x_1 и x_2 может принимать различные значения при заданном значении произведения ($x_1 x_2$). В том случае, когда имеет место взаимное влияние переменных, осуществлять масштабирование гораздо сложнее. Для исключения членов, содержащих произведение двух переменных, квадратичные формы можно привести к каноническому виду. При этом определяются новые координатные оси, называемые *главными осями,* относительно которых данная поверхность второго порядка симметрична и содержит только аргументы в квадрате. В новой системе координат масштабирование каждого члена значительно проще, чем в исходной системе.

Нелинейные неквадратичные функции можно сделать квадратичными при помощи подходящего преобразования модели или разложения соответствующих функций в ряд Тейлора с сохранением конечного числа членов.

5. В математической модели может оказаться, что некоторые из переменных исключаются путем надлежащего преобразования. Пусть, например,

$$y = x_1^2 + 2x_1x_2 + x_2^2 + 10 = (x_1 + x_2)^2 + 10.$$

После преобразования $x_1 + x_2 = \tilde{x}$, получим $y = \tilde{x}^2 + 10.$ Таким образом, вместо двух переменных в выражении для y осталась только одна переменная $\tilde{x}$, которую и следует варьировать, чтобы найти экстремум функции y.

4. Задачи математического программирования.

Как уже отмечалось во введении, сейчас общий класс математического программирования в зависимости от вида целевой функции и ограничений распадается на линейное программирование, целочисленное программирование, динамическое программирование, выпуклое программирование, нелинейное программирование и программирование при наличии неопределенности.

Для большинства экономических и технических задач наиболее часто применяются методы линейного и нелинейного программирования. Рассмотрим их формулировки и особенности.

4.1 Задача линейного программирования.

Задачей линейного программирования называется задача, в которой минимизируемым или максимизируемым критерием является *линейная* функция, причем на переменные налагаются ограничения, которые также представляются *линейными* функциями. Комбинация скаляров или векторов, обычно обозначаемых X_i, называется *линейной,* если она может быть записана в виде

$$c_1X_1 + c_2X_2 + \ldots + c_nX_n\,, \tag{1}$$

где все коэффициенты c — константы. Например, функция

$$4\,x_1 + 3\,x_2 + 5\,x_3 + 2$$

линейна относительно переменных x_1, x_2, x_3, тогда как функция

$$2x_1^2 + x_1x_2 + 3e^{x_3}$$

не линейна относительно тех же переменных. Величины X_1 ..., X_n называются *линейно зависимыми,* если для некоторого набора c_i в предположении, что не все c_i равны нулю, имеет место следующее равенство:

$$c_1X_1 + c_2X_2 + ... + c_nX_n = \sum_{i=1}^{n} c_iX_i = 0. \qquad (2)$$

С другой стороны, если $\sum_{i=1}^{n} c_iX_i = 0$ только тогда, когда все коэффициенты c_i равны нулю, то величины X_i ... X_n называются *линейно независимыми.*

Линейное программирование стало быстро развиваться после второй мировой войны, привлекая внимание математиков, экономистов и инженеров благодаря возможности широкого практического применения, а также математической стройности. Применение линейного программирования оказалось плодотворным в таких областях, как:

1. Оптимальное регулирование полетов на воздушных линиях.

2. Распределение во времени транспортировки грузов с заводов и складов к базам.

4. Планирование промышленного производства.

5. Система оплаты контрактов.

6. Проектирование систем связи и распределение в них потоков сообщений.

7. Распределение кадров.

8. Организация максимальных потоков в транспортных сетях.

Широкое привлечение внимания к линейному программированию привело к тому, что до некоторой степени была преувеличена его значимость; часто упускали из виду, что оно применимо лишь тогда, когда выполняются лежащие в его основе допущения, которые в свою очередь базируются на *линейном* математическом представлении реального мира. В нелинейном программировании не используются столь сильные упрощения.

Хотя задача линейного программирования может быть сформулирована по-разному, запишем ее в следующей форме:

минимизировать $$y = \sum_{i=1}^{n} c_i x_i \Rightarrow \min \qquad (3)$$

при ограничениях

$$\sum_{i=1}^{n} a_{ij} x_i - b_j \ge 0, \qquad j = 1,2,\ldots m; \qquad (4)$$

$$x_i \ge 0, \qquad i = 1,2,\ldots m; \qquad (5)$$

где a, b и c — константы, а x_i — искомые переменные.

Уравнения (3)-(5) можно компактно записать в матричной форме. Пусть X и **c** – вектор-столбцы размерности (n × 1); **a** – матрица констант размерности (n × m) и **b** – вектор-столбец размерности (m × 1):

$$X = \begin{Bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{Bmatrix}, \ \mathrm{a} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nm} \end{bmatrix}, \ \mathrm{b} = \begin{Bmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{Bmatrix}, \ c = \begin{Bmatrix} c_1 \\ c_2 \\ \ldots \\ c_n \end{Bmatrix}.$$

Тогда в матричных обозначениях уравнения (3)-(5) запишутся следующим образом:

минимизировать $$y = f(X) = c^T X \Rightarrow \min \qquad (6)$$

при ограничениях

$$a^T X \geq b; \qquad (7)$$

$$X \geq 0; \qquad (8)$$

где верхний индекс T означает транспонирование.

Для решения задачи, выраженной уравнениями (6)-(8), были предложены различные методы, один из которых (симплекс-метод) стал общеупотребительным и который, в частности, кратко рассмотрен в учебном пособии по методам оптимизации [3].

Пример. Этот пример иллюстрирует связь между уравнениями (3)-(5) и реальной задачей. Крупная авиакомпания ассигновала 600 млн. долл. на лизинг самолётов трех типов. Годовой лизинг самолёта *A* стоит 10 млн. долл., самолёта *B* — 20 млн. долл. и самолёта *C* — 23 млн. долл. Сколько самолётов каждого типа нужно заказать, чтобы получить наибольшую производительность в тонно-километрах на один день с учетом приведенных ниже условий?

Самолёт *A* требует для эксплуатации одного экипажа на каждую смену, максимальное число смен в сутки 3, производительность 21000 тонно-километров за каждую смену.

Самолёт *B* требует двух экипажей на каждую смену, максимальное число смен в сутки 3, производительность 36000 тонно-километров за каждую смену.

Самолёт *C* требует двух экипажей на каждую смену, максимальное число смен в сутки 3, производительность 37800 тонно-километров за каждую смену.

Кроме того, число экипажей должно быть не больше 145, число самолётов — не больше 30.

Эта задача может быть сформулирована в терминах линейного программирования следующим образом. Обозначим число смен индексами 1, 2 и 3, а число самолётов каждого типа — прописными буквами A, B и C. Таким образом A_1 означает количество самолётов типа A, работающих одну смену в сутки.

Сначала необходимо выбрать функцию цели. В данном примере критерием оптимальности является производительность, соответствующая функции (3):

$y = 21000A_1 + 42000A_2 + 63000A_3 + 36000B_1 + 72000B_2 + 108000B_3 + 37800C_1 + 75600C_2 + 113400C_3,$

где y - количество тонно-километров. Запишем далее выражения для ограничений. Здесь мы имеем дело с ограничениями трех типов: 1) общая стоимость лизинга всех самолётов не должна превышать 600 млн. долл.; 2) количество самолётов должно быть не более 30; 3) количество экипажей должно быть не более 145:

1) 10 млн. $(A_1 + A_2 + A_3)$ + 20 млн.$(B_1 + B_2 + B_3)$ + 23 млн. $(C_1 + C_2 + C_3) \leq$ 600 млн.;

2) $A_1 + A_2 + A_2 + B_1 + B_2 + B_3 + C_l + C_2 + C_3 \leq 30$;

3) $A_1 + 2A_2 + 3A_3 + 2B_1 + 4B_2 + 6B_3 + 2C_1 + 4C_2 + 6C_3 \leq 145$.

Кроме того, поскольку самолёты являются физическими объектами, то их количество не может быть отрицательной величиной, и тогда имеют место следующие неравенства, соответствующие условиям (5):

$A_i \geq 0, B_i \geq 0, C_i \geq 0, i = 1, 2, 3.$

Поставленная таким образом задача может быть решена соответствующим методом линейного программирования; в результате определяется количество самолётов типов A, B и C, максимизирующее величину целевой функции (производительность) y.

Решение: **12** самолётов типа *A*, **0** самолётов типа *B* и **18** самолётов типа *C*.

4.2. Общая задача нелинейного программирования.

В самом широком смысле общая задача нелинейного программирования заключается в отыскании экстремума целевой функции при заданных ограничениях в виде равенств и/или неравенств. Ограничения могут быть линейными и/или нелинейными. Однако общепринята несколько более узкая постановка общей задачи нелинейного программирования, в которой исключаются из рассмотрения следующие специальные случаи:

1. Переменные принимают лишь целочисленные значения (нелинейное целочисленное программирование).

2. Ограничения включают в качестве параметра время, при этом используются дифференциальные уравнения (оптимальное управление, динамическая оптимизация).

Пусть непрерывная функция $f(\mathbf{x})$ представляет собой целевую функцию; $h_1(X),\ldots\ h_m(X)$ задают ограничения в виде равенств, а $g_{m+1}(X),\ldots g_p(X)$ — ограничения в виде неравенств, где $X = \{x_1,\ldots\ x_n\}^T$ - вектор-столбец компонентов $x_1,\ldots x_n$ в n-мерном евклидовом пространстве E^n.

Как и в линейном программировании, переменные $x_1,\ldots x_n$ могут быть конструктивными параметрами, установками регулятора, показаниями измерительных приборов и т.д., тогда как целевая функция может представлять собой стоимость, вес, годовой доход и т.п., а ограничения - технические требования, условия работы, пропускную способность или факторы безопасности, присущие данному процессу.

Формально задача нелинейного программирования может быть сформулирована следующим образом:

минимизировать $$f(X) \Rightarrow \min, \quad X \in E^n, \qquad (9)$$

при m линейных и/или нелинейных ограничениях в виде равенств

$$h_j(X)=0, \qquad j=1,2,...m; \tag{10}$$

и *(p - m)* линейных и (или) нелинейных ограничениях в виде неравенств

$$g_j(X)\geq 0, \qquad j=m+1,m+2,...p; \tag{11}$$

Хотя в некоторых частных случаях ограничения в виде равенств могут быть явно разрешены относительно выбранных переменных, которые затем исключаются из задачи как независимые переменные, и в задаче остаются только ограничения в виде неравенств, чаще всего ограничения в виде равенств не могут быть явно разрешены и потому сохраняются.

Иногда встречается другое представление выражений (9) - (11):

минимизировать $\{f(X)\,|\,X\in R\}\Rightarrow \min;$ (12)

где R - область пространства переменной X, для которой выполнены условия (10) и (11), например:

$$R=\{\,X\,|\,h_j(X)=0,\;\; g_j(X)\geq 0 \text{ для всех } \;\; j\,\}. \tag{13}$$

В выражениях (12) и (13) знак | читается как «при», а символ $\in$ означает «принадлежит». Знак неравенства в выражении $g_j(X)\geq 0$ может быть изменен на обратный путем умножения на (-1), что не изменит математической постановки задачи.

В качестве простого примера задачи нелинейного программирования, которую можно проиллюстрировать графически, приведем следующую:

минимизировать $f(X) = x_1^2 + x_2^2 + 2x_2$

при ограничениях

$h_1(X) = x_1^2 + x_2^2 - 1 = 0;$

$g_2(X) = x_1 + 2x_2 - 0.5 \geq 0;$

$g_3(X) = x_1 \geq 0;$

$g_4(X) = x_2 \geq 0.$

На рис. 5 целевая функция изображена пунктирными линиями, ограничение в виде равенства - жирной сплошной кривой, а граница допустимой области, определяемой неравенствами, - линиями со штриховкой, нанесённой с внутренней стороны области.

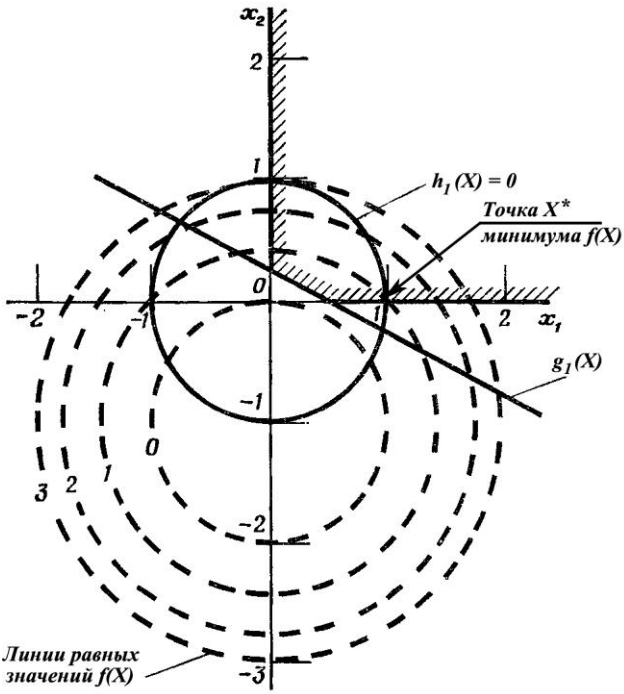


Рис.5. Топография целевой функции и ограничений

В каждой точке n-мерного пространства переменных [x_1, x_2,... x_n] функция $f(X)$ имеет определённое значение, и, следовательно, это n-мерное пространство представляет собой скалярное поле для критерия оптимальности. Как показано на рис.5, в этом пространстве можно вычертить семейство линий равных значений (эквипотенциальных гиперповерхностей) для выбранных значений функции $f(X)$.

Пространство переменных $[x_1, x_2, \ldots x_n]$ является также скалярным полем для функций ограничений, для которых также можно изобразить графически эквипотенциальные гиперповерхности. При помощи классических методов анализа в общем случае невозможно заранее выявить расположение точки X^*, которая соответствует минимальному (или максимальному) значению функции $f(X)$, поскольку она может находиться на пересечении ограничивающих поверхностей или между ними.

Задачи линейного и *квадратичного программирования* могут рассматриваться как два частных случая общей задачи нелинейного программирования. Если и функция (9), и уравнения (10), и неравенства (11) линейны, то мы имеем дело с задачей линейного программирования. Если же целевая функция квадратичная, а ограничения линейные, то имеет место задача квадратичного программирования:

минимизировать

$$f(X) = a_0 + c^T X + X^T \mathrm{Q} X \Rightarrow \min \tag{14}$$

при ограничениях

$$a^T X \geq b; \tag{15}$$

$$X \geq 0; \tag{16}$$

где **Q** - положительно определенная или неотрицательно определенная квадратная симметрическая матрица; **a** и **b** - матрицы коэффициентов, определение которых дано ранее в связи с уравнением (7). Иногда в задачу квадратичного программирования включают линейные ограничения в виде равенств

$$a^T X = b. \tag{17}$$

Во всех трех классах задач - нелинейных, линейных и квадратичных - нужно найти вектор $X^* = \{x_1, x_2,... x_n\}^T$, минимизирующий или, наоборот, максимизирующий функцию $f(X)$ при следующих условиях: $h_j(X) = 0,\ j = 1,2,...m;$ и $g_j(X) \geq 0,\ j = m+1, m+2,...p;$

4.3. Основные понятия теории оптимизации.

4.3.1. Оптимальные решения.

Вектор $X^* = \{x_1, x_2,... x_n\}^T$, удовлетворяющий соотношениям (10) - (11), называется *оптимальной точкой,* а соответствующее значение $f(X^*)$ - оптимальным значением целевой функции. Пара X^* и $f(X^*)$ составляет *оптимальное решение.* Как показано на примере мультимодальной (многоэкстремальной) функции на рис. 6, могут существовать различные типы оптимальных решений, если целевая функция не является *унимодальной* (т. е. имеющей один экстремум). Глобальное оптимальное решение представляет собой наименьшее значение $f(X)$, тогда как *локальное* (или относительное) оптимальное решение представляет собой наименьшее значение $f(X)$ в окрестности некоторого вектора X. Как для глобального, так и для локального минимума

$$f(X^*) \leq f(X),$$

но для глобального оптимального решения это соотношение выполняется для всех X в эвклидовом пространстве E^n, тогда как для локального оптимального решения это имеет место только для малой области ξ, где норма $\| X - X^* \| < \xi$. Если принимается во внимание и точность решения, то условие оптимальности можно представить в виде

$$f(X^*) \leq f(X) - \gamma,$$

где γ — некоторая малая величина.

Все алгоритмы, описываемые в пособии [3], дают лишь локально оптимальные решения, так как на каждом этапе решения при движении к точке экстремума X^* они зависят в основном от локальных свойств (местной топографии) целевой функции и ограничений. Работа всех поисковых алгоритмов примерно одинакова: на основании исследования топографии целевой функции в текущей точке определяется характер её изменения и в пространстве проектных переменных делается рабочий шаг в направлении её уменьшения (или увеличения). Далее, процесс повторяется. Исключение составляют чистые методы Монте-Карло, которые исследуемые точки выбирают случайным образом. Однако эти методы вообще не гарантируют ничего: экстремум они находят случайно, если вообще находят.

На практике предположение о том, что локальный экстремум является глобальным, может быть проверено путем использования нескольких начальных векторов, но даже если найдено только одно локальное решение, в общем случае нельзя показать, что это решение обязательно является глобальным оптимумом.

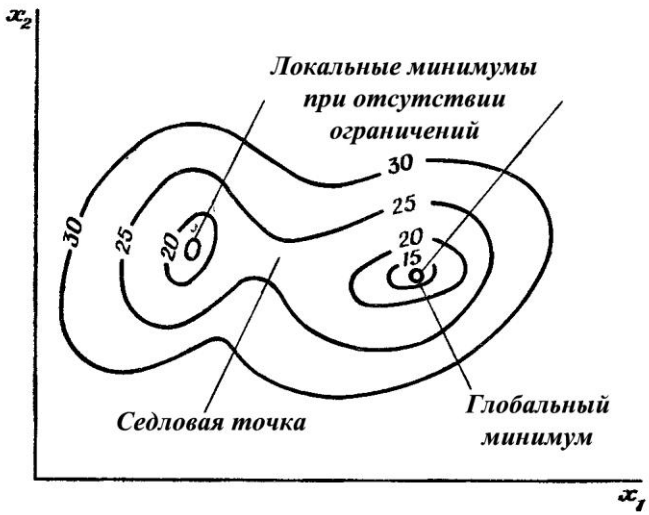


Рис. 6. Классификация оптимальных решений.

К счастью, для задач, соответствующих действительным физическим процессам, целевая функция обычно является хорошей и обладает единственным экстремумом. Поэтому для большинства практических целей использование численных процедур, дающих локальное решение задачи оптимизации, не является большим недостатком.

4.3.2. Вогнутость и выпуклость.

Понятия вогнутости и выпуклости полезны при определении того, при каких условиях локальное оптимальное решение является также глобальным оптимальным решением, что представляется важным в случае множественных оптимумов.

Функция $\Phi(X)$ называется *выпуклой* в области R, если для *любых* двух векторов X_1 и $X_2 \in R$ выполняется неравенство

$$\Phi(\theta X_1 + (1-\theta) X_2) \le \theta \Phi(X_1) + (1-\theta)\Phi(X_2); \qquad (18)$$

где θ - скалярная величина из диапазона $0 \le \theta \le 1$. Функция $\Phi(X)$ является *строго выпуклой,* если для $X_1 \ne X_2$ в выражении (18) используется знак неравенства (<). Векторное неравенство $X > Y$ означает, что $x_i > y_i$ для каждого элемента; в случае $X > Y$ справедливо неравенство $x_i > y_i$ для всех i. На рис. 7 показаны строго выпуклая и строго вогнутая функции одной переменной. Выпуклая функция $\Phi(X)$ не может принимать значение, большее, чем значение, полученное линейной интерполяцией в промежутке между X_1 и X_2. Если имеет место неравенство, обратное (18), то функция называется *вогнутой*.

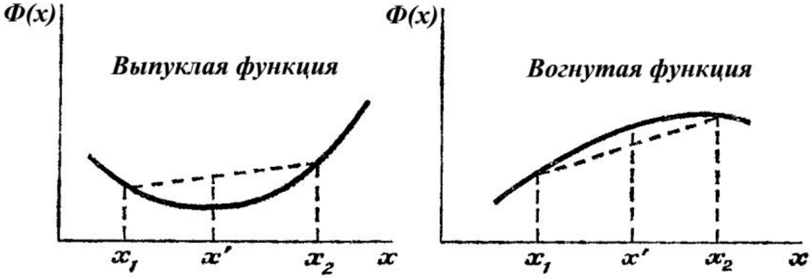


Рис. 7.

Линейные функции одновременно и выпуклые и вогнутые, так как неравенство (18) выполняется для любых точек как равенство.

Дифференцируемая выпуклая функция обладает следующими свойствами.

1. $$\Phi(X_2)-\Phi(X_1)\geq\begin{Bmatrix}\dfrac{\partial\Phi}{\partial x_1^1}\\ \dfrac{\partial\Phi}{\partial x_2^1}\\ \ldots\\ \dfrac{\partial\Phi}{\partial x_n^1}\end{Bmatrix}^T(X_2-X_1)$$ для любых X_1 и X_2. Другая запись этого условия такова: $\Phi(X_2)-\Phi(X_1)\geq\nabla^T\Phi(X_1)\cdot(X_2-X_1)$ для любых X_1 и X_2.

2. Матрица вторых частных производных функции $\Phi(X)$ по вектору X, так называемая матрица Гессе, положительно определённая (или положительно полуопределённая) для всех X, если $\Phi(X)$ строго выпуклая (или просто выпуклая). Если определитель матрицы Гессе равен нулю, то матрица Гессе может считаться как положительно, так и отрицательно полуопределённой.

3. В области допустимых значений R функция $\Phi(X)$ имеет только один экстремум, то есть вогнутая функция унимодальна.

Множество точек или область в n-мерном пространстве называется выпуклым, если для любых пар точек X_1 и X_2, принадлежащих этому множеству, отрезок прямой, соединяющий их, также полностью принадлежит множеству, см. рис. 8.

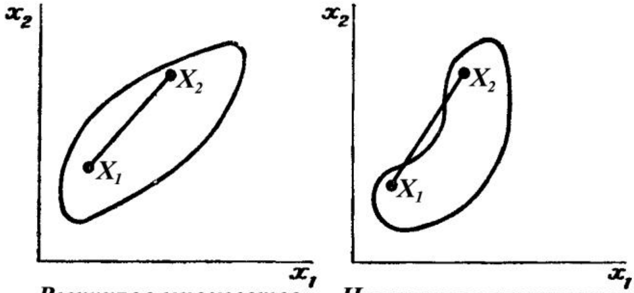


Рис. 8.

Из понятия выпуклости вытекает один важный результат математического программирования. Задача нелинейного программирования, представленная в виде *задачи выпуклого программирования*, формулируется следующим образом:

минимизировать $f(X) \Rightarrow \min$

при ограничениях

$$g_j(X) \geq 0, \quad j = 1,2,...p;$$

$$X \geq 0;$$

где $f(X)$ – *выпуклая* функция и каждая функция $g_j(X)$ - *вогнутая* функция. При этом локальный минимум является также и глобальным минимумом. Аналогично локальный максимум является глобальным максимумом, если целевая функция вогнутая, а ограничения образуют выпуклое множество.

Рассмотрим следующую задачу (рис. 9): минимизировать $f(X) = x_1^2 + x_2^2 - 4x_1 + 4$;

при ограничениях:

$g_1(X) = x_1 - x_2 + 2 \geq 0;$

$g_2(X) = -x_1^2 + x_2 - 1 \geq 0;$

$g_3(X) = x_1 \geq 0;$

$g_4(X) = x_2 \geq 0.$

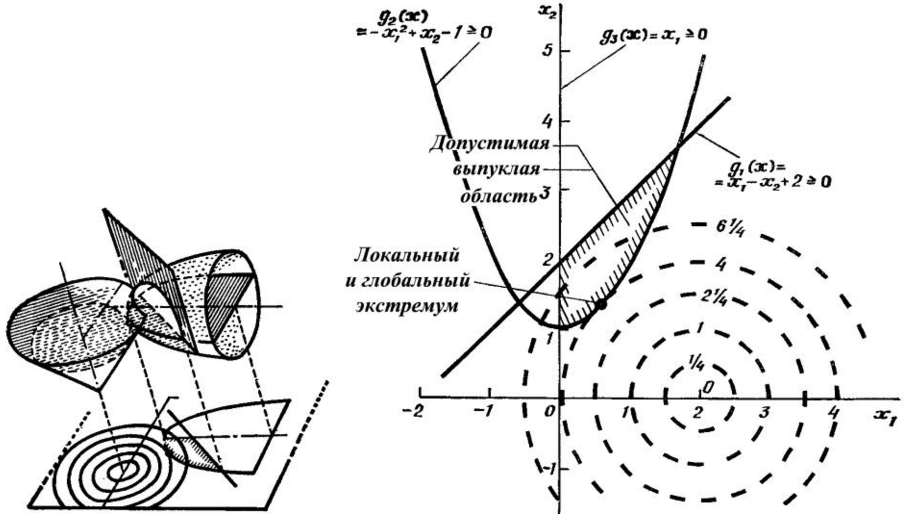


Рис. 9. Задача выпуклого программирования.

Внимательное изучение кривых (к тому же выводу можно прийти и аналитически) показывает, что ограничения образуют выпуклую область (допустимая часть области заштрихована), поскольку функции $g_1\ (X)$, $g_3\ (X)$ и $g_4\ (X)$ линейные и, следовательно, вогнутые (правда, они одновременно и выпуклые), а функция $g_2\ (X)$ строго вогнутая. Можно показать, что функция g_2(x) вогнутая, заметив, что матрица Гессе функции ($-g_2(x)$) положительно полуопределённая.

$$-\begin{bmatrix} \dfrac{\partial^2 g_2(X)}{\partial x_1^2} & \dfrac{\partial^2 g_2(X)}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 g_2(X)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 g_2(X)}{\partial x_2^2} \end{bmatrix} = -\begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix} = 0.$$

Целевая функция $f(X)$ строго выпуклая, и локальный оптимум, являющийся также и глобальным оптимумом, достигается в точке $X^* = \{0.58\ 1.34\}^T$, где $f(X) = 3.80$; см. рис. 9.

В задаче *линейного программирования*, имеющей оптимальное решение, целевая функция всегда выпуклая, а ограничения образуют выпуклое множество, так что локальный оптимум всегда является в то же время и глобальным оптимумом.

Ограничения в задаче *квадратичного программирования* те же, что и в задаче линейного программирования, а целевая функция выпуклая, если (X^T **Q** X) является положительно полуопределённой; следовательно, матрица **Q** должна быть неотрицательно определенной. Однако только при особых обстоятельствах можно показать, что общая нелинейная функция $f(X)$ является выпуклой, а ограничения образуют выпуклое множество. Так, одной из самых серьезных трудностей является то, что *нелинейные равенства* не могут быть частью выпуклой области, содержащей больше чем одну точку, поскольку прямая линия, соединяющая две любые несмежные точки, удовлетворяющие данному равенству, не может в то же самое время проходить и через другие точки, удовлетворяющие этому равенству, как это требуется для выпуклости. Тем не менее в особом случае, когда ограничивающее множество образовано лишь ограничениями в виде неравенств и все функции-ограничения являются вогнутыми, так что точки, для которых $g_j(X) \geq 0$, образуют выпуклое множество, задача *нелинейного программирования* может представлять собой задачу выпуклого программирования.

4.3.3. Допустимость.

Любой вектор X, удовлетворяющий ограничениям в виде равенств и в виде неравенств, называется *допустимой точкой* или допустимым вектором. Множество всех точек, удовлетворяющих ограничениям, образует допустимую область функции $f(X)$, которую будем обозначать через R; любая точка вне R называется *недопустимой. Условный оптимум* представляет собой локальный оптимум, лежащий на границе допустимой области. Если ограничения имеют вид равенств, то допустимый вектор X должен лежать на пересечении всех гиперповерхностей, соответствующих

$h_j(X) = 0$. При ограничениях в виде неравенств точка X может быть либо *внутренней точкой* (допустимой точкой), либо *граничной точкой* (тоже допустимой точкой), либо *внешней точкой* (недопустимой точкой). Для внутренней точки $g_j(X) > 0$; в случае граничной точки удовлетворяется по крайней мере одно равенство $g_j(X) = 0$, а для внешней точки имеет место по крайней мере одно неравенство $g_j(X) < 0$. Множество точек, удовлетворяющих равенствам $g_j(X) = 0$, $j = 1,2...p$, определяет *граничные поверхности* системы ограничений, заданных в виде неравенств. *Активным,* или *связывающим,* ограничивающим неравенством называется такое, которое для данного вектора X превращается в равенство $g_j(X) = 0$. Область допустимых значений переменных может быть односвязной (фиг. 10*а)* и многосвязной (фиг. 10б). В последнем случае может оказаться, что конкретный алгоритм нелинейного программирования проведёт обследование не всех допустимых областей, если процедура поиска не будет включать большое число начальных векторов.

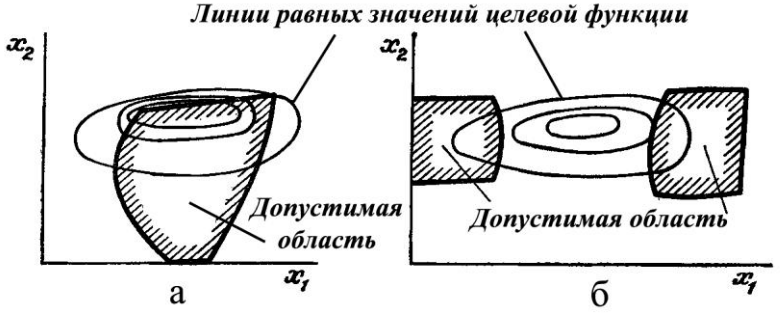


Рис. 10. Примеры односвязной (а) и многосвязной (б) допустимой области

К счастью, большинство задач нелинейного программирования, относящихся к реальным процессам, формулируются так, что допустимая область является односвязной.

4.3.4. Градиент.

Множество точек, для которых целевая функция имеет одинаковое значение, называется линией равных значений или *линией уровня f(X)*. Несколько таких линий уровней изображено на рис. 11.

Если целевая функция непрерывна и дифференцируема, то существует *градиент f(X)*, определяемый как *вектор-столбец* из первых частных производных *f(X)* по *X*, значения которых берутся в данной точке *X*. Верхний индекс *k*, *k* = 0,1..., используется для обозначения точки в евклидовом пространстве E^n, в которой берется значение градиента, и, таким образом, градиент в точке $X^{(k)}$ равен

$$\nabla f(X^{(k)}) = \begin{Bmatrix} \dfrac{\partial f(X^{(k)})}{\partial x_1} \\ \dots \\ \dfrac{\partial f(X^{(k)})}{\partial x_n} \end{Bmatrix}.$$

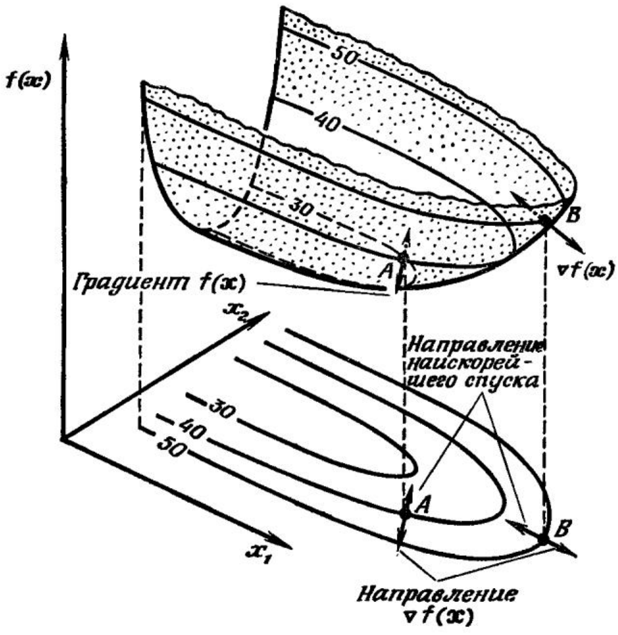


Рис. 11. Градиент и направление наискорейшего спуска.

Выражение $\nabla^T f(X^{(k)})$ означает *вектор-строку*. В матричной алгебре доказывается, что градиент скалярной функции направлен в сторону наискорейшего увеличения функции, то есть *наискорейшего подъема,* и что он ортогонален линии уровня функции $f(X)$, проходящей через данную точку $X^{(k)}$. Вектор, противоположный этому градиенту (отрицательный градиент), направлен в сторону *наискорейшего спуска.* Любой вектор $\boldsymbol{V}$, ортогональный $\nabla f(X^{(k)})$ [так же, как и касательная плоскость к $f(X^{(k)})$ в точке $X^{(k)}$], может быть записан как имеющий следующее свойство: $V^T \nabla f(X^{(k)}) = 0$.

Следует отметить, что градиент не является направлением наискорейшего увеличения $f(X)$, если рассмотрение ведется не в евклидовой, а в какой-то другой метрике. Например, если определять длину вектора X не по формуле $|| X || = (X^T X)^{1/2}$, а по формуле $\| X \| = \sum_j |x_j|$, то направление наискорейшего увеличения $f(X)$ можно получить, находя ту компоненту $\nabla f(X^{(k)})$, которая имеет наибольшее абсолютное значение, и полагая соответствующую компоненту X равной либо (+1), либо (-1) в зависимости от знака компоненты градиента, а остальные компоненты приравнивая нулю, как, например, в симплекс-методе при линейном программировании.

4.3.5. Аппроксимация функций.

Для некоторых процедур математического программирования необходимо осуществлять линейную или квадратичную аппроксимацию функций $f(X)$, $g_j(X)$ и $h_j(X)$. Например, линейная, или первого порядка, аппроксимация целевой функции $f(X)$ может быть выполнена с помощью усеченного ряда Тейлора в окрестности $X^{(k)}$:

$$f(X) \approx f(X^{(k)}) + \nabla^T f(X^{(k)}) \cdot (X - X^{(k)}). \qquad (19)$$

Квадратичную аппроксимацию функции *f(X)* можно получить, отбрасывая в рядах Тейлора члены третьего и более высокого порядка:

$$f(X) \approx f(X^{(k)}) + \nabla^T f(X^{(k)}) \cdot (X - X^{(k)}) +$$

$$+\frac{1}{2}(X - X^{(k)})^T \cdot \nabla^2 f(X^{(k)}) \cdot (X - X^{(k)}), \quad (20)$$

где $\nabla^2 f(X^{(k)})$ - *матрица Гессе* функции *f(X)*, которая обозначается **H**(*X*) и представляет собой квадратную матрицу вторых частных производных *f(X)*, взятых в точке $X^{(k)}$:

$$\nabla^2 f(X^{(k)}) = H(X^{(k)}) = \begin{bmatrix} \frac{\partial^2 f(X^{(k)})}{\partial x_1^2} & \ldots & \frac{\partial^2 f(X^{(k)})}{\partial x_1 \partial x_n} \\ \ldots & \ldots & \ldots \\ \frac{\partial^2 f(X^{(k)})}{\partial x_n \partial x_1} & \ldots & \frac{\partial^2 f(X^{(k)})}{\partial x_n^2} \end{bmatrix}. \quad (21)$$

5. Необходимые и достаточные условия оптимальности.

В области нелинейного программирования большое внимание было уделено определению необходимых и достаточных условий того, чтобы некоторый вектор решения *X** являлся локальным экстремумом. Критерий оптимальности решения для некоторых особых случаев нелинейного программирования, описанных формулами (9) – (11), был сформулирован, когда мы обсуждали выпуклость целевой функции и ограничений. Однако структура задачи нелинейного программирования в общем случае такова, что полностью критерий оптимальности еще не разработан. Вследствие этого рассмотрим только случай оптимизации функций *без ограничений* и в дальнейшем покажем приёмы сведения задач с ограничениями к задачам без ограничений.

Алгоритмы поиска экстремума, работающие непосредственно с ограничениями, требуют больших вычислительных затрат, тонкой настройки в конкретных практических задачах и, на наш взгляд, представляют интерес в основном для специалистов по вычислительной математике. Инженерам надо решать задачи, а не анализировать тонкости работы изящных, но малопродуктивных поисковых алгоритмов.

5.1. Нелинейное программирование без ограничений.

Постановка задачи:

Минимизировать $f(X), \quad X \in E^n.$ (22)

Для задачи нелинейного программирования при отсутствии ограничений *необходимыми условиями* того, что *X** — точка локального минимума задачи (22), являются:

1) функция *f(X)* дифференцируема в точке *X**;

2) $\nabla f(X^*) = 0$, то есть в точке *X** имеет место стационарная точка.

Достаточным условием того, что *X** - точка локального минимума задачи (22), кроме приведенных условий 1) и 2), дополнительно является следующее:

3) $\nabla^2 f(X^*) > 0$, то есть матрица Гессе положительно определённая.

Соответствующие условия для максимума те же самые, за исключением того, что матрица Гессе для *f(X*)* должна быть отрицательно определённой.

Однако, возможно существование минимума (максимума), который не удовлетворяет необходимым и достаточным условиям, например, см. рис. 12, но если достаточные условия удовлетворены, то *X** обязательно будет точкой минимума.

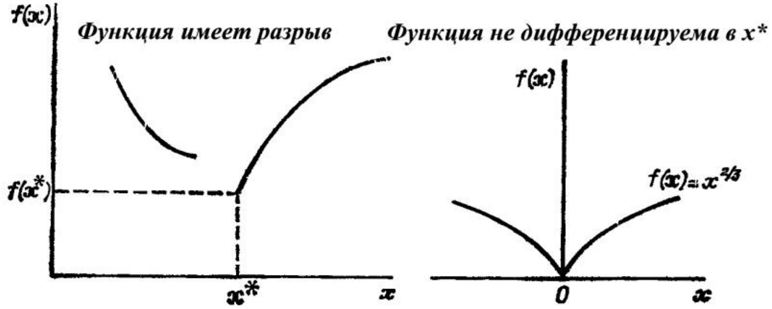


Рис. 12. Возможные точки экстремума.

6. Сведение задачи с ограничениями к безусловно экстремальной форме.

Поисковые методы отыскания экстремума в задачах с явным учётом ограничений являются инструментом очень тонким и расточительным в смысле вычислительных ресурсов. С другой стороны, методы, разработанные для безусловной оптимизации, обладают высоким быстродействием, требуют относительно небольшой памяти и позволяют решать практические задачи в автоматическом режиме без каких-либо настроек алгоритмов [3].

В этой связи заманчиво преобразовать исходную задачу с ограничениями в эквивалентную задачу без ограничений, причём такое преобразование должно обеспечивать совпадение точек экстремума исходной и преобразованной задачи. Основным методом сведения задачи условной оптимизации к безусловно экстремальной форме является *метод штрафных функций*.

Идея метода проста: если мы ищем минимум целевой функции и, если текущая точка поиска находится внутри допустимой области, то штрафная добавка к значению целевой функции не прибавляется; если же какое-либо ограничение нарушено, то к целевой функции прибавляется штрафная добавка, пропорциональная степени нарушения ограничения. В случае поиска максимума штрафная добавка должна вычитаться.

Это заставляет поисковый алгоритм работать внутри допустимой области и при попытке выхода из неё, возвращаться обратно.

Методы штрафных функций существуют в различных вариантах, которые, отличаются способом формирования штрафной добавки, однако, имеют одну общую черту: во всех этих методах осуществляется преобразование задачи нелинейного программирования при наличии ограничений либо в одну (эквивалентную исходной) задачу без ограничений, либо в эквивалентную последовательность задач без ограничений. Пусть, например, мы хотим найти минимум функции

$$f(X) = (x_1 - 3)^2 + (x_2 - 2)^2$$

при ограничении

$$h(X) = x_1 + x_2 - 4 = 0.$$

Прибавим $h^2(X)$ к целевой функции $f(X)$, в результате чего получим новую целевую функцию

$$P(X) = (x_1 - 3)^2 + (x_2 - 2)^2 + (x_1 + x_2 - 4)^2,$$

значения которой будем считать свободными от каких-либо ограничений ($h^2(X)$ здесь выступает в роли «штрафа»). В процессе минимизации $P(X)$ «штрафная добавка» к $f(X)$ способствует тому, чтобы вектор X в некоторой степени удовлетворял исходному ограничивающему условию $h(X) = 0$. Совершенно очевидно, что, пока условие $h(X) = 0$ удовлетворяется (в пределах установленного допуска), значение штрафного члена пренебрежимо мало, и $P(X^*) \to f(X^*)$ при $X \to X^*$. Преимущество, которое мы получаем за счет перехода от задачи минимизации при наличии ограничений к задаче минимизации в отсутствие ограничений, состоит в том, что в последнем случае минимизация может осуществляться с помощью гораздо более простых (по сравнению с первым случаем) алгоритмов. При использовании методов штрафных функций получается максимальный оптимизирующий эффект за счет постоянного компромисса между необходимостью удовлетворения ограничений и процессом минимизации $f(X)$, который достигается путем присвоения надлежащих весов целевой функции и функциям, задающим ограничения.

На рис. 13 графически представлена преобразованная целевая функция, полученная путем добавления к *f(X)* штрафной функции (с весовым коэффициентом, равным единице). Отметим, что точка минимума X^* у обеих функций одинакова.

Методы штрафных функций можно разделить на два класса: 1) параметрические и 2) непараметрические методы.

Непараметрические методы рассматривают целевую функцию как функцию, задающую дополнительное искусственное ограничение, постепенно «уплотняемое» по мере получения информации в ходе поиска. Непараметрические методы, на наш взгляд, это просто игра ума специалистов по вычислительной математике, результаты которой иногда, к их удивлению, позволяет решить какую-то практическую задачу.

Нам нужны надёжные автоматические методы, поэтому мы рассмотрим только параметрические методы штрафных функций. Название этих методов говорит о том, что в структуру штрафной функции, которая строится с помощью функций-ограничений, входят подобранные параметры в качестве весовых коэффициентов.

Параметрические методы распадаются на три категории: 1) методы внутреннего штрафа; 2) методы внешнего штрафа и 3) комбинированные методы.

При использовании методов внутреннего штрафа значение целевой функции удерживается в отдалении от границы допустимой области, то есть точка $X^{(k)}$ постоянно находится внутри допустимой области с помощью штрафной функции. Методы внешней точки, наоборот, генерируют последовательность точек, которые выходят за пределы допустимой области, но дают в пределе допустимое решение. Штрафная функция не позволяет вектору *X* слишком удаляться от границы допустимой области. В комбинированных методах, использование которых особенно необходимо в случае, когда явно учитываются ограничения в виде

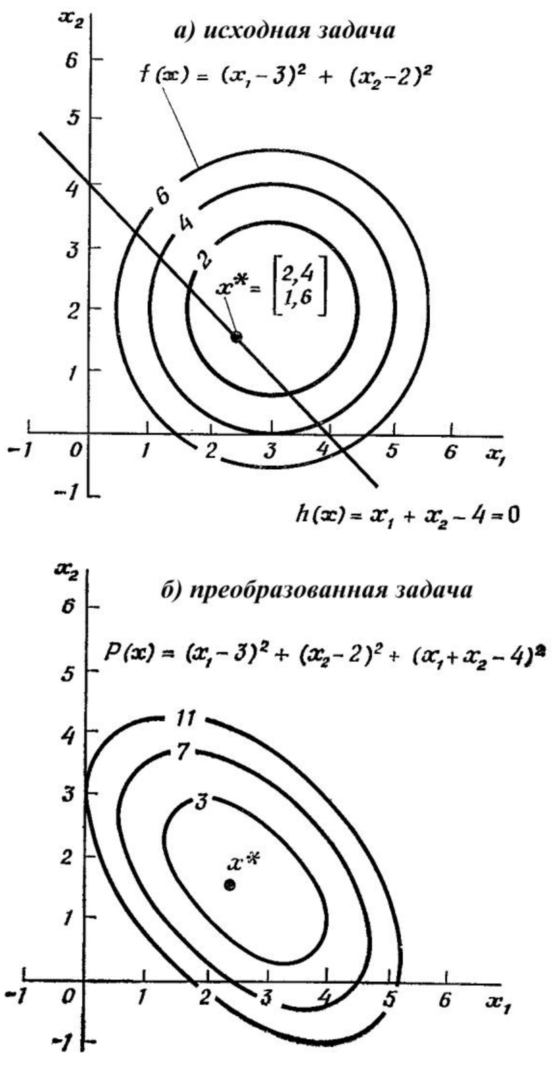


Рис. 13. Линии уровней исходной а) и преобразованной б) задачи.

равенств, в ходе минимизации одни из ограничивающих условий удовлетворяются, а другие не удовлетворяются; однако все условия в пределах заданного допуска оказываются удовлетворенными при достижении оптимального решения.

Формулировка большинства технических задач содержит только ограничения-неравенства; в организационно-экономических задачах часто имеют место ограничения-равенства, например, нужно оптимальным образом потратить определённый ресурс, - однако равенства легко преобразуются в ограничения-неравенства путем задания небольшого допуска:

$$g_j(X) = \varepsilon_j - \left| h_j(X) \right| \geq 0, \tag{23}$$

где ε_j - величина допуска нарушения ограничения $h_j(X) = 0$. С учетом этого, будем считать, что имеем дело только с ограничениями-неравенствами, то есть формулировку задачи нелинейного программирования (9), (10), (11) сокращаем, модифицируем и применяем в виде, характерным для технических задач:

$$\text{минимизировать } f(X) \Rightarrow \min, \quad X \in E^n, \tag{24}$$

при p линейных и (или) нелинейных ограничениях в виде неравенств

$$g_j(X) \leq 0, \qquad j = 1, 2, \ldots p; \tag{25}$$

6.1. Метод внешнего штрафа.

При использовании внешнего штрафа расширенная функция выглядит следующим образом:

$$P(X) = f(X) + \sum_{j=1}^{p} \begin{cases} 0, \text{ } если \;\; g_j(X) \leq 0; \\ k_j \, g_j(X), \text{ } если \;\; g_j(X) > 0; \end{cases}. \tag{26}$$

На рис. 14 показано как формируется расширенная (штрафная) функция для функции одной переменной и одном геометрическом ограничении $x \leq a$.

Видно, что в методе внешнего штрафа поисковый алгоритм для активного ограничения в качестве оптимальной всегда будет находить точку, лежащую почти на границе допустимой области, но снаружи неё. В этом его основной недостаток. К достоинствам следует отнести то, что начальная точка может лежать где угодно: штрафная добавка всегда вернет текущую точку внутрь допустимой области.

Степень выхода оптимальной точки за границу допустимой области зависит от величины весовых коэффициентов k_j в формуле (26). Они определяют приемлемую степень перехода за границу допустимой области для каждого отдельного ограничения и должны подбираться в каждой конкретной задаче.

Для технических задач, связанных с ограничениями по прочности, жёсткости, динамическим характеристикам и т.п. приемлемую точность дают коэффициенты $k_j = 10^4$. Это очень сильное требование и в этом случае компьютерная программа реализации поискового алгоритма должна оперировать числами с двойной точностью (64 бита или 16 десятичных цифр мантиссы).

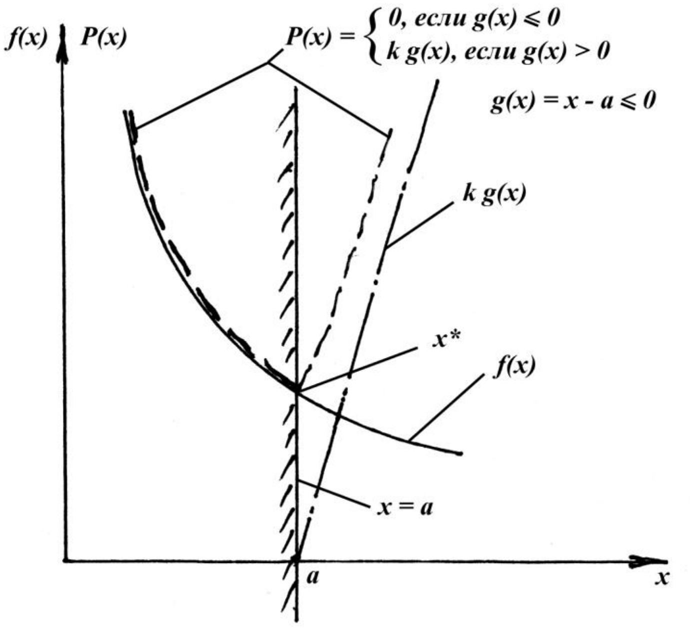


Рис. 14. Внешняя штрафная функция

6.2. Метод внутреннего штрафа.

Идея внутреннего штрафа состоит в том, что штрафная добавка не должна позволять поисковому алгоритму не то, чтобы переходить границу допустимой области, но даже приближаться к ней. Поэтому методы внутреннего штрафа называют ещё методами барьерных функций. Расширенная функция формируется в следующем виде:

$$P(X) = f(X) + \frac{1}{\sum_{j=1}^{p} \left| g_j(X) \right|}. \quad (27)$$

На рис. 15 показана расширенная функция одной переменной с использованием внутреннего штрафа. Видно, что точка оптимума гарантированно лежит внутри допустимой области.

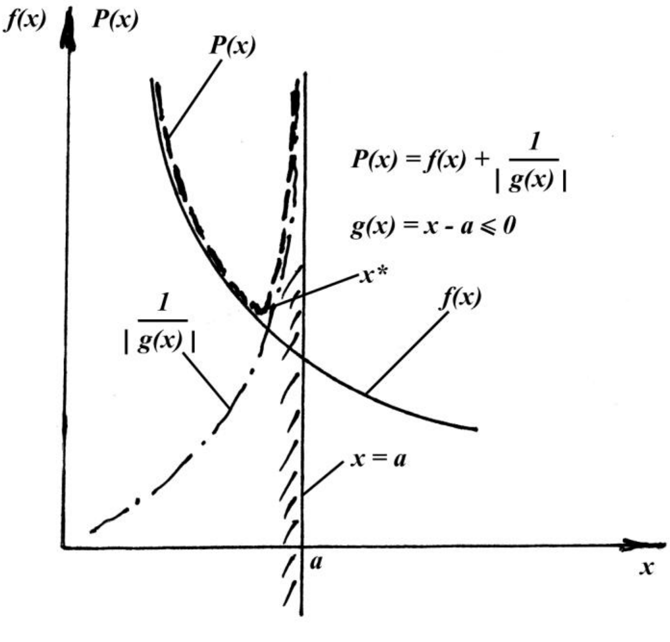


Рис. 15. Метод внутреннего штрафа.

При этом, однако, всегда существует некоторый «зазор», отделяющий найденный экстремум от принципиально достижимого оптимума. Убрать этот зазор довольно просто: после оптимизации нужно задать соответствующий параметр, лежащим на границе допустимой области и вычислить целевую функцию; однако регулировать его в процессе поиска довольно затруднительно.

Вторым недостатком метода внутреннего штрафа являет то, что барьер на границе допустимой области работает при приближении к границе как изнутри области, так и снаружи неё. Поэтому, если в процессе поиска текущая точка вышла за пределы допустимой области, то обратно она не вернётся никогда. Это обстоятельство необходимо учитывать при задании начальной точки поиска: она обязательно должна лежать внутри допустимой области. Метод внутреннего штрафа хорошо работает на монотонных функциях, в противном случае он может приводить к ложным экстремумам.

6.3. Комбинированный метод штрафных функций.

Является объединением методов внешнего и внутреннего штрафа. В этом методе все ограничения разбиваются на две группы: для одной группы ограничений применяется внешний штраф, для другой – внутренний. Расширенная функция при этом имеет вид:

$$P(X) = f(X) + \frac{1}{\sum_{j=1}^{m} \left| g_j(X) \right|} + \sum_{j=m+1}^{p} \begin{cases} 0, \; если \;\; g_j(X) \le 0; \\ k_j \, g_j(X), \; если \;\; g_j(X) > 0. \end{cases} \tag{28}$$

Методы штрафных функций любого типа избавляют нас от явного учёта ограничений, однако привносят некоторую трудность в выбор и программную реализацию поисковых алгоритмов. Дело в том, что штрафные добавки портят топографию исходной целевой функции, добавляя в неё изогнутые «овраги», которые имеют настолько крутые склоны, что их можно считать каньонами.

На рис. 16 показан пример такой овражной функции, которая в настоящий момент является общепризнанным тестом для анализа скорости и качества работы поисковых алгоритмов. Алгоритм в тестах запускается из начальной точки $X = \{-1.2, 1\}^T$ и должен найти точку минимума $X^* = \{1, 1\}^T$.

Применяемые в методах штрафных функций поисковые алгоритмы должны уметь не только спускаться в овраг, но и «ходить» по его изогнутому дну в поисках экстремума. Отметим, что градиентные методы этого не умеют: они спускаются в овраг и останавливаются, считая, что минимум найден. В пособии [3] специально рассмотрены именно овражные методы.

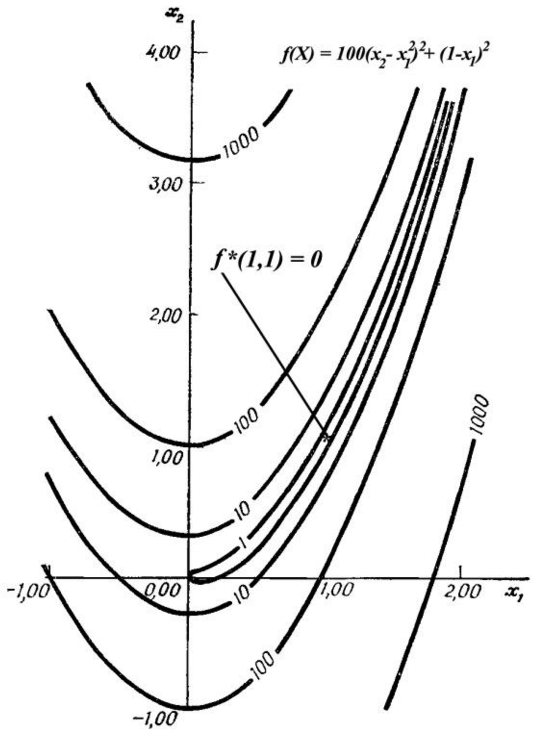


Рис. 16. Линии равных значений функции Розенброка.

6.4. Метод множителей Лагранжа.

Этот метод широко используется при оптимизации авиационных агрегатов при их проектировании с учётом многообразных и разнородных ограничений, например: проектирование конструкции с учётом ограничений по прочности, жёсткости, частотам собственных колебаний, устойчивости и т.п. Поэтому краткое знакомство на уровне идей с методом множителей Лагранжа представляется полезным.

Методы, основанные на использовании множителей Лагранжа, относятся к категории параметрических методов штрафных функций, поскольку для них характерно то, что функции-ограничения вводятся в структуру модифицированной целевой функции совместно с некоторым переменным параметром. Чтобы обобщить метод множителей Лагранжа, ограничения в виде неравенств следует преобразовать в ограничения, имеющие вид равенств, путём введения надлежащих ослабляющих переменных (на каждое ограничение-неравенство по одной ослабляющей переменной). Задача нелинейного программирования в общей постановке принимает при этом следующий вид:

$$\text{минимизировать } f(X),\ X \in E^n; \qquad (29)$$

при ограничениях

$$h_j(X) = 0;\quad j = 1,2 \ldots m; \qquad (30)$$

$$g_j(X) - v_j^2 = 0;\quad j = (m+1) \ldots p. \qquad (31)$$

Если вычесть v_j^2 из $g_j(X)$ ($j = m + 1, \ldots p$), то можно гарантировать, что ограничивающее условие, имеющее в исходной постановке задачи вид неравенства, действительно выполняется. Тогда можно определить обычным образом функцию Лагранжа

$$L(X,\Lambda) = f(X) + \sum_{j=1}^{m} \lambda_j h_j(X) + \sum_{j=m+1}^{p} \lambda_j \left[g_j(X) - v_j^2 \right]; \qquad (32)$$

где λ_j, $j = 1,2...p$; - неотрицательные и не зависящие от *X* весовые коэффициенты, которые можно отождествить с множителями Лагранжа. Для того чтобы *X** было решением общей задачи нелинейного программирования (29)-(31), необходимо и достаточно, чтобы: 1) функция *f(X*)* была выпуклой; 2) в окрестности *X** ограничения задачи были выпуклы и 3) в точке *X** удовлетворялась следующая система уравнений, определяющая стационарное решение уравнения (32):

$$\frac{\partial L}{\partial x_i} = 0; \qquad \text{при} \quad i = 1, 2...n;$$

$$\frac{\partial L}{\partial \lambda j} = 0; \qquad \text{при} \quad j = 1, 2 \ldots p;$$

$$\frac{\partial L}{\partial v j} = 2\lambda_j v_j = 0; \qquad \text{при} \quad j = (m+1)\ldots p;$$

$$\lambda_j \geq 0; \qquad j = 1,2\ldots p.$$

Короче говоря, условный минимум *f(X)* имеет место в стационарной точке для *L(X, Λ, V)* и, в частности, в седловой точке (*X, Λ, V*) - пространства, так что задача с ограничениями превращается в задачу определения седловой точки в отсутствие ограничений.

7. Заключение.

На этом мы завершаем краткое рассмотрение общей теории оптимизации. Конечно, мы не исчерпали весь круг проблем, необходимых для решения при оптимизации, однако теперь читатель сможет уверенно и компетентно формулировать задачи оптимизации организационных и технических систем, классифицировать их и выбирать для их решения наиболее эффективные поисковые алгоритмы, большой набор которых приведен в книге [4].

За рамками пособия осталась большая область теории, связанной с многопараметрической оптимизацией, с инженерными методами оптимизации, а также с оптимизацией в условиях неопределённости. Поэтому данная тонкая книжка должна рассматриваться как введение в большой и плодотворный мир оптимального проектирования организационно-технических систем.

Литература

1. Уайлд Д. Оптимальное проектирование. -М: Мир, 1981. -272с.

2. Батищев Д.И. Методы оптимального проектирования. М.: Радио и связь, 1984

3. Данилин А.И. Методы оптимизации. Учебное пособие. – Самара: Изд-во Самарского государственного аэрокосмического ун-та, 2011. -66с.

4. Химмельблау Д. Прикладное нелинейное программирование. М: Мир, 1975. -534с.

Оглавление

1. Введение.. 3

2. Постановка задачи оптимизации.. 8

2.1. Функция цели.. 9

2.2. Проектные переменные... 11

2.3. Запись ограничений... 14

3. Построение модели.. 15

3.1. Принятие допущений... 16

3.1.1. Некоторые часто используемые допущения......................... 17

3.2. Использование физических принципов.................................... 19

3.3. Вычислительные трудности... 20

4. Задачи математического программирования............................. 23

4.1 Задача линейного программирования....................................... 23

4.2. Общая задача нелинейного программирования........................ 28

4.3. Основные понятия теории оптимизации................................... 32

4.3.1. Оптимальные решения... 32

4.3.2. Вогнутость и выпуклость... 34

4.3.3. Допустимость... 38

4.3.4. Градиент... 39

4.3.5. Аппроксимация функций.. 41

5. Необходимые и достаточные условия оптимальности............. 42

5.1. Нелинейное программирование без ограничений...................... 43

6. Сведение задачи с ограничениями к безусловно экстремальной форме... 44

6.1. Метод внешнего штрафа. 48

6.2. Метод внутреннего штрафа. 50

6.3. Комбинированный метод штрафных функций. 51

6.4. Метод множителей Лагранжа. 53

7. Заключение. 55

Литература 55